

It's a New Year Requiring Immediate Access to New Data With Continued Instant Access to Prior Data

It's a new year. But that doesn't mean starting from a clean slate from a text search perspective. Rather, a new year requires immediate access to new data with continued instant access to prior years' data. It's a tall order, but enterprise search is up to the task.

Enterprise search like dtSearch® can work in multiple different instant concurrent-access search capacities. First, enterprise search can run on a standard Windows network. Second, enterprise search can run from an in-house web-based intranet server. Third, enterprise search can run from the cloud such as AWS or Azure.

The network option has some advantages for those in-office, while a web-based and potentially cloud-based approach has benefits for those operating remotely. Whether network-based, intranet-local server or cloud-based, enterprise search offers instant concurrent searching only after first indexing the data. All installations can share the same indexes, making it possible to easily run enterprise search in parallel configurations if that makes sense for the organization.

With dtSearch, each index can hold up to a terabyte of text and there are no limits on the number of indexes the software can create and simultaneously search. One or more indexes can hold previous years' data, with fresh indexes holding data for the new year. For dynamic data, use the Windows Task Scheduler to automatically update indexes to reflect new, modified and deleted content. Updating an index does not block out current searching. While indexing takes a lot of system resources, searching does not. Immediate multithreaded

Article contributed
by dtSearch®



The network option has some advantages for those in-office, while a web-based and potentially cloud-based approach has benefits for those operating remotely. Whether network-based, intranet-local server or cloud-based, enterprise search offers instant concurrent searching only after first indexing the data.

indexed searching can therefore proceed even while indexes continuously update.

To start indexing, just tell dtSearch the email archives, folders and the like you want the index to cover, and the indexer will take it from there. To correctly parse each file, enterprise search needs to determine the file format of each item. But file extensions can be misleading, such as an email archive that someone saved with a .PDF extension. For accurate file format detection, dtSearch bypasses the file extension entirely and looks directly at the binary file format to determine file type.

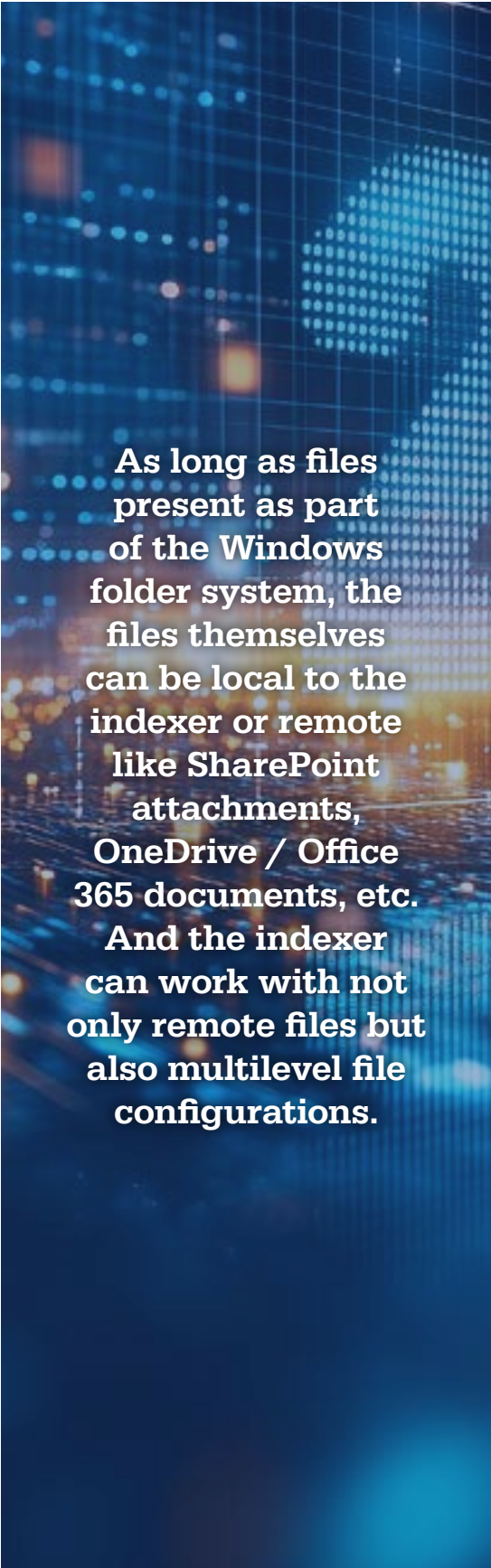
As long as files present as part of the Windows folder system, the files themselves can be local to the indexer or remote like SharePoint attachments, OneDrive / Office 365 documents, etc. And the indexer can work with not only remote files but also multilevel file configurations. For example, the data can include an email with a ZIP or RAR attachment that has a PDF Portfolio plus a Word document which itself embeds a spreadsheet and dtSearch will still index all text and metadata.

After indexing, the software has more than 25 different search features to enable sifting through data. An “all words” search request for *Project Codename Alpha Zeta* would look for only files that contain all of these search terms. An “any words” search request for *Project Codename Alpha Zeta* would look for files that contain even just one of these search terms. An “exact phrase” search request would look for *Project Codename Alpha Zeta* as a precise phrase match.

By default, searching covers all text plus metadata. But you can limit a search to specific metadata or positionally to the top or bottom of a file. A Boolean search request enables more intricate search formulations like the phrase *Project Codename Alpha Zeta* in a document that also has *Texas* or *Arizona* but not *New Mexico*. Or use proximity search to find *Project Codename Alpha Zeta* within a certain number of words of a different phrase like budgetary planning.

Particularly relevant for multiyear data, searching can also include date or date range components like *date(November 15, 2023 to January 27, 2024)*. This date range search would also pick up date variants like *Jan 5*

Article contributed
by dtSearch®



As long as files present as part of the Windows folder system, the files themselves can be local to the indexer or remote like SharePoint attachments, OneDrive / Office 365 documents, etc. And the indexer can work with not only remote files but also multilevel file configurations.

2024 and 1/17/24. Like other search components, a date can be anywhere in indexed files. Alternatively, a date search can look only in specific metadata.

The software can also look for number or numeric ranges and even for some number patterns like identifying any credit card numbers across indexed data. Concept searching extends a search for project to undertaking. Stemming finds different endings on the same root word like projects, projecting and projection in a search for project. Fuzzy searching sifts through typographical and OCR errors, like *projext* for *project*.

And searching isn't limited to English. dtSearch supports Unicode spanning hundreds of international languages. A single file can cycle through multiple languages, including not only different European languages but also right-to-left Hebrew and Arabic and double-byte Chinese, Japanese and Korean. dtSearch and Unicode will follow that whole progression.

For ordering search results, dtSearch uses vector-space relevancy ranking. Take any "any words" query for *Project Codename Alpha Zeta*. If project, codename and alpha are common across indexed data while zeta rare, then files containing zeta will get a higher relevancy rank, with files with the densest zeta mentions coming out highest. Or end-users can define their own term weightings, like giving codename a positive weight of 3, *alpha* a negative weight of 3 and *zeta* a positive weight of 9 but only if it occurs in certain metadata or at the top or bottom of a file. For a different view on search results, end-users can instantly re-sort by some other metric like filename or file data. Whatever the sorting, the software can display retrieved files with highlighted hits for easy review.

Article contributed
by dtSearch®



dtSearch supports Unicode spanning hundreds of international languages. A single file can cycle through multiple languages, including not only different European languages but also right-to-left Hebrew and Arabic and double-byte Chinese, Japanese and Korean. dtSearch and Unicode will follow that whole progression.